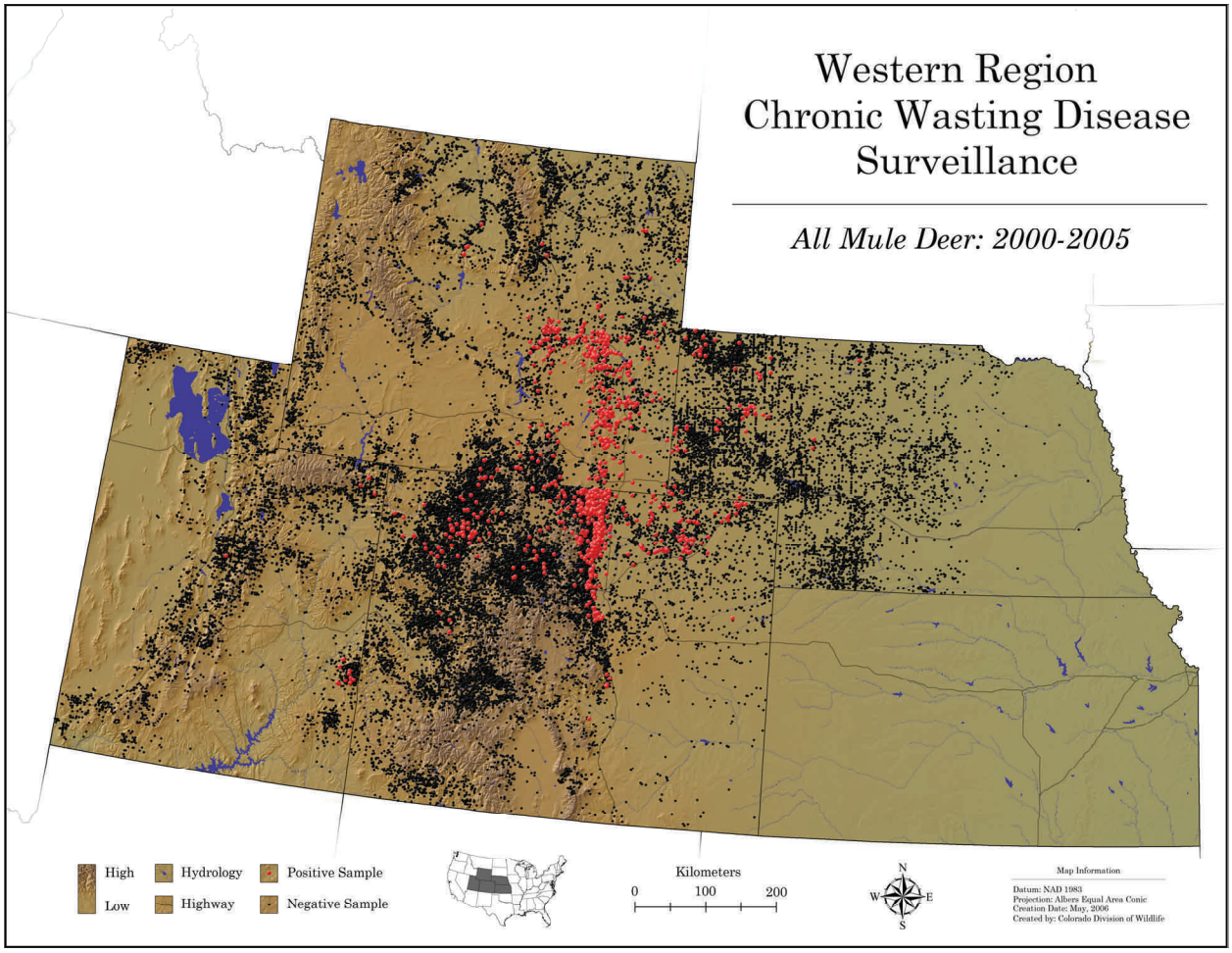
Approaches to Regional-Scale Modeling



Chronic wasting disease samples from the Western United States region (2000-2005).

We consider the term regional-scale to describe large areas; large enough to be relevant to multi-jurisdictional (e.g., multi-state or multi-providence) management efforts and regional description of CWD epidemiology. Regional-scale epidemiology focuses on the biotic and abiotic factors that influence the observed spatial pattern of a disease, and its spread between populations (see Hess et al. 2002) or across large areas. At the regional scale, data are generally summarized by county, wildlife management unit, or other relatively large areas, such as a state or providence. Data collected typically include time to next case (rate) or counts of positive and negative samples within a given area (area is also called "tract" in epidemiology literature). For a region, major goals include evaluating biological and ecological risk factors and predicting high-risk areas. At this scale, researchers and managers also attempt to identify likely corridors of disease spread and potential barriers that could be used to arrest proliferation.

From a regional perspective, an introduced wildlife disease may appear as a point source with diffusion (Figure 1.1), as observed for bovine tuberculosis in white-tailed deer (Schmitt et al. 1997, Hickling 2002) in Michigan, USA, and raccoon rabies in the northeast-

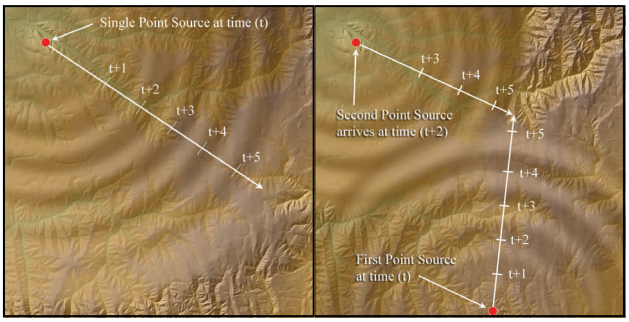


Figure 1.1. *Visual depiction of wave front phenomena emanating from single (left) and multiple sources (right).*

ern USA (Jenkins and Winkler 1987, Moore 1999). Established epidemics may show a diffusion wave front, as seen in fox rabies in Europe (Kallèn et al. 1985, Smith and Harris 1991), or multiple point sources, as seen in anthrax epidemics in African ecosystems (Prins and Weyerhaeuser 1987). Although a regional view of an epidemic may suggest diffusion across a landscape, finer resolution may reveal a more patchy distribution and heterogeneity in the rate of spread. For example, the pattern of raccoon rabies in Pennsylvania, USA, appeared consistent with simple diffusion when viewed from a large, geographic perspective. Subsequent analyses, however, revealed areas of slow spread (barriers), high prevalence areas, and rapid local spread (corridors) that did not conform to simple diffusion model predictions (Moore 1999). Predictions at the regional-scale can be useful to describe disease occurrence, prevalence, or spread in general terms or for political reasons. Yet, patterns of disease spread at a biologically relevant scale may be poorly represented by values averaged over a large spatial area.

Spatial epidemiological models on a large scale are geared toward one of two distinct forms of disease: highly contagious infectious diseases that generate rapidly moving epidemic fronts, such as foot-and-mouth disease, or non-infectious diseases, such as non-viral cancers or toxin related illness. Because of the chronic nature of CWD infection in animals, the putative slow rate of transmission within populations, and the relatively slow rate of geographic spread, CWD epidemiology falls somewhere between these two rubrics and thus we describe approaches for both types of disease.

RISK ANALYSIS / ASSESSMENT:

Risk analysis is a term often used when evaluating effects of a known point-source (such as a power plant) or line-source (such as contaminated streams or rivers) that emits toxins or pollutants relative to proximity to the source (Morris and Wakefield 2000). Risk analysis, or risk assessment, also describes geographical studies in terms of spatial distribution of putative risk factors, and their relationship or correlation to disease risk (Briggs 2000). Overall or total risk is the response variable, which can be a variety of functions, usually summarized in a map. For example, risk could be a krigged surface of prevalence or probability of infection, or a relative risk surface or odds-ratio surface. Risk can also be expressed qualitatively, where areas are described as having high, medium, or low risk. The risk response surface could be course, wherein all risks are averaged over a cell, or continuous, such as for krigged data. The choice of a risk response variable and its spatial resolution depends on study goals and available data.

Diseases have multiple causes, and disentangling the risk factors can be complex. However, this problem is analogous to other spatial predictive models and analyses are typically conducted in a GIS framework, in which environmental or ecological landscape characteristics are used to predict presence (MacKenzie et al. 2006), distribution (Murwira and Skidmore 2005), abundance (Dunford and Freemark 2005), or other spatial traits, such as home range size (Anderson et al. 2006) of wildlife species. Spatial data have frequently been used to link predictions about species distribution and abundance to habitat characteristics, but it is less common to use this approach in predicting disease distribution or prevalence.

Perhaps this is because regionally-oriented risk models are not applicable to highly infectious, fast-spreading diseases, which have been the historical focus of spatial epidemiological modeling for regional scales. Risk analyses are typically more appropriate for endemic diseases that are spreading slowly, or not at all. In the case of a slowly spreading disease, one wants to be reasonably sure that absence of the disease is related to factors associated with the site (or individual) rather than it simply being farther away from the site where the disease was introduced. Note, however, that in more sophisticated analyses, distance from the epicenter and time since introduction could be included as covariates prior to investigating the effects of other variables. Several groups have employed spatial risk models to predict the probability of disease, or disease-vector presence, based on environmental factors: tsetse flies (Rogers et al. 1996), Lyme disease (Allen et al. 2003, Schauber et al. 2005), human induced disease of great apes (Sleeman 2005), and disease-carrying *Ixodex ricinus* ticks (Merler et al. 1996).

Questions addressed / model predictions:

1. Estimates the relationship between environmental/ecological (abiotic and biotic) factors and disease risk.
2. Potentially estimates mechanistic relationships between disease and environmental/ ecological (abiotic and biotic) factors.
3.

Data required:

1. If disease cases present themselves, such as cancer cases, then only spatial locations of the positive samples are required (note that "targeted" surveillance samples for CWD may fall in this category).
2. Spatial coordinates of positive and negative samples are required if disease cases do not 'present' themselves – such as for CWD cases.
3. If any temporal aspect to the models, then the date samples were collected (as a proxy for date of infection of individuals.
4. Spatial environmental/ecological (abiotic and biotic) data for factors of interest.

Output:

1. Estimates spatial variation in disease prevalence and/or risk.
2. Estimates risk parameters, effect sizes, covariate effect sizes, and other relevant statistics for factors/variables affecting the probability of disease.
3. Provides model selection statistics and the relative weight of different models/ variables.

General usefulness:

Spatial risk models are potentially useful for non-infectious disease or diseases with low infection rates or slow epidemic fronts (i.e., relatively slow spread). Because wildlife hosts are mobile, this approach is most applicable for diseases with short dormancy or latency periods. Longer latencies would dilute the effect of some factors unless the risk factors were relatively constant and migration rates were relatively low (e.g., animals do not contact disease in one location and then move to another so that risk is unhinged from location of infected animals), whereas short latency diseases can be more directly tied to risk factors. Thus, ecological risk factors associated with short latency period are more readily identified. Spatial risk models could be used to test various hypotheses about environmental and biological covariates and risk factors, as well as to identify potential areas of disease risk.

Usefulness to CWD modeling and/or management:

If relevant spatial data are available, risk analysis has high potential applicability to CWD modeling and management. CWD could make use of a risk analysis approach, given the fact that geographical spread appears to be slow in nature (we find it where we look for it, and, at present, have seen no evidence of rapid or even moderate rate of spatial spread). Surveillance data would satisfy the three data requirements, but availability of appropriate environmental and ecological spatial data would need to be assessed, collected, and evaluated. This approach could identify environmental/ ecological risk factors to potentially target for

management, as well as high prevalence and high risk areas to target management intervention or other actions. For example, a recent study found that prions bind with varying affinity to different soil minerals (Johnson et al. 2006) and Farnsworth et al. (2005) found greater CWD risks in urban areas. Risk analyses could be used to predict how soil types and housing density may affect current and future hotspots of CWD, but the power of the analyses would depend upon host movement, spatial scale of variation in soil type and the amount and resolution of available data.

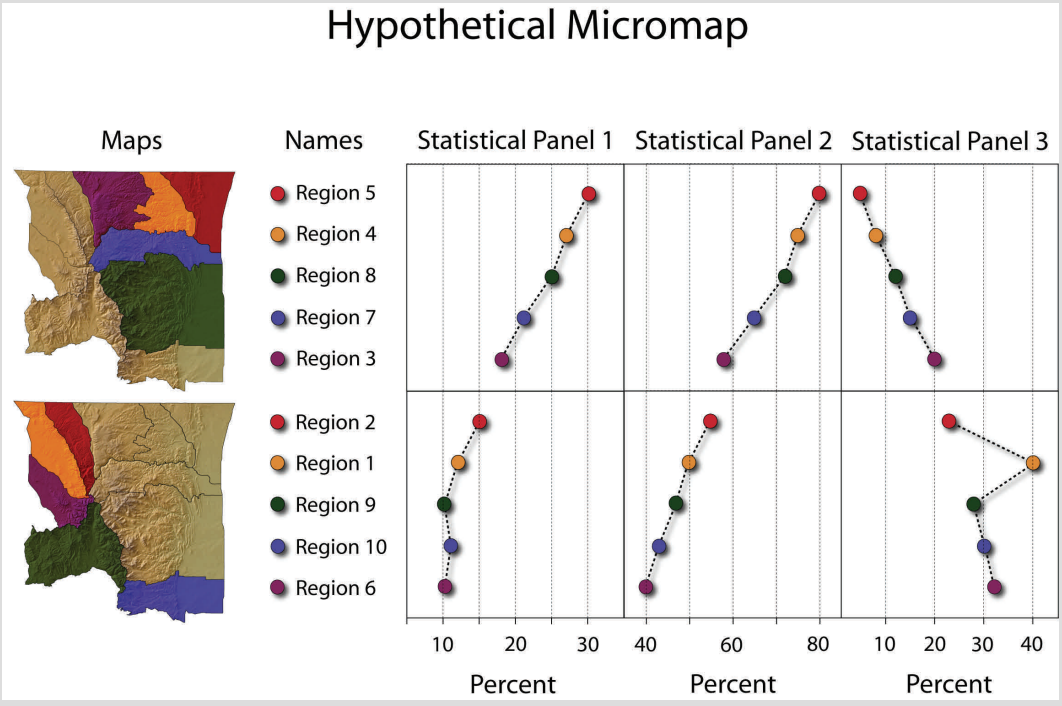
MICROMAPS

Micromaps represent a different paradigm for graphical visualization of diseases (Box 1.1) compared to other common techniques such as choropleth maps (see Figure 2.6) or smoothed prevalence or risk surfaces (see Figure 1.7). The chloropleth map, which uses either shading or color to differentiate the values from one region to another, is somewhat problematic as noted by Dent (1993) and Harris (1999). The main issue is the color representation of continuous data, which usually necessitates the conversion of continuous data into discrete intervals and going from infinite color gradation to a limited number of colors (Symanzik and Carr 2007). Symanzik and Carr (2007) provide details on issues pertaining to region area size, the loss of information when making continuous information discrete and the inability to represent confounding variables.

Given that most CWD data is georeferenced, and the response variables of primary interest are numerical, linked micromap plots can be used to order and present this multivariate data in a contextual structure. In addition, confidence intervals for variables, such as prevalence, can be shown on micromaps. This is an important piece of information missing from chloropleth maps. Thus, linked micro-

Box 1.1. A CLOSER LOOK AT MICROMAPS

A typical, but hypothetical, linked micromap plot showing the four key features of the technique (Carr and Pierson 1996). Note: maps display simulated (not real) data. The first (leftmost) panel, Maps, contains a map of the region. The second panel, ID, provides the names of the geographical regions (here, Region 1 through Region 10). The third through the fifth panels display statistical summaries. These panels may represent many forms of statistical summaries including boxplots, dot-plots (as shown), time series plots, confidence intervals, etc. Sorting the geographic regions based on the statistical variable(s) of interest is the second feature. Sorting improves perception between consecutive panels from top to bottom of the display. The



Box 1.1. A CLOSER LOOK AT MICROMAPS *(continued from last page)*

third feature is the partitioning of the regions into perceptual groups of size five or less to allow the viewer's attention to focus on explicit areas at a time. The fourth feature is color and location that links corresponding elements within the parallel sequence panels, i.e., the color red in the topmost panels relates to the geographic region in the Northeast of the map, the area name Region 5, and a red dot in each of the three statistical panels. The color red is reused in the next consecutive set of panels for Region 2, but there is no relationship between Region 5 and Region 2 as one might at first assume. Simply, there are not enough distinguishable colors to populate an entire display (with, say, 50 different regions); consequently, colors have to be reused in different panels.

The data displayed in statistical panels 1 and 2 show a strong positive association (the correlation r calculated as 0.99), expressed in the almost parallel behavior of the dots and lines representing the values for these two variables. In contrast, the statistical data in panel 3 and 1 (as well as 3 and 2) show a strong negative association (the correlation r calculated as –0.94 for 3 and 1 and as –0.92 for 3 and 2). This negative association is seen in the movement of the dots and lines in opposite directions for these variables. Moreover, the data in panel 3 show an unusual outlier, the value for Region 1. It is this outlier that considerably reduces the almost perfect negative association otherwise present in this data. Just a simple numerical calculation of r might not be able to reveal the influence of a single region on the overall relationship.

The map panels of the linked micromap plots from the previous page exhibit a strong geographic pattern: Highest occurrences with respect to the statistical panel 1 can be found in the North and in the East; lowest occurrences can be found in the West and in the South. Additional features of linked micromap plots exist and are described in more detail in Symanzik and Carr (2007).

maps plots are visual, but also are a representation of statistical data without certain shortcomings of the chloropleth depiction (Carr 2001). This provides benefits through (a) more meaningful representation of the data and, (b) communicating complex data sets in a manner that facilitates interpretation.

An ample set of templates are available that offer users considerable flexibility in data visualization (Carr et al. 1998). For example, the statistical panel of linked micromap plots can use any number of forms including box-plots, bar-plots, histogram-plots, or time series plots (Box 1.1). These alternate statistical plots offer additional avenues to represent the underling structure of the data and examine patterns and relationships in the data.

Questions addressed / model predictions:

1. Depicts prevalence or other relevant disease statistics or information across space and time simultaneously.
2. Allows queries of the underling structure of the data.
3. Facilitates examination of patterns and relationships in the data.

Data required:

1. Summary of positive and negative samples by polygon, such as wildlife management unit, county, etc., for each time step of interest.
2. The polygon from which each positive and negative sample was taken.

Output:

1. Prevalence maps that are linked across space and time simultaneously.
2. Any statistics of interest linked across space and time simultaneously.

General usefulness:

Linked micromap plots facilitate data sort-

ing and division into smaller groups, which can be used to highlight specific spatial and temporal patterns. Linked micromap plots provide considerably more information than would otherwise be provided by a series of tables or an overall map representation (e.g., a chloropleth map) alone. Viewers can navigate through the linked micromap plot to a place of interest in order to review prevalence and related statistics, such as confidence intervals. Thus, estimates and the equally important variance of those estimates can be portrayed. This is an advantage over typical spatial summaries wherein variance is not portrayed. A second key advantage is that linked micromaps present statistical summaries and estimates in a spatial context. Unlike traditional graphical methods, linked micromap plots combine both exploratory analysis and traditional statistical graphics while maintaining the spatial context; this is important in CWD epidemiology because of its intrinsic spatial nature.

Usefulness to CWD modeling and/or management:

Linked micromaps provide highly useful visualization techniques for presenting temporal data from a large number of areas, such as wildlife management units or counties, along with associated statistics. These plots can allay information overload and facilitate interpretation of large and complex data sets; this property is extremely useful to managers who need to make timely decisions about CWD management. Linked micromap plots are a constructive GIS representation coupled to a statistical visualization tool, which provide exploratory capabilities. Linked micromap plots can be used to augment the presentation of CWD data. One example may be to micromap the results of a risk assessment model that would stratify areas based on risk to facilitate planning and mitigation practices.

CLUSTER ANALYSIS

Cluster analysis refers to a widely used set of grouping algorithms that identify meaningful structures (often spatial for spatial disease epidemiology) in observed data. The conceptual approach involves grouping data so that patterns within a valid cluster are more similar to each other than to patterns belonging to different clusters (Jain et al, 1999). The scope of problems addressed by cluster analysis includes many disciplines and has led to the development of a large assortment of clustering methods. No single clustering technique is universally appropriate for uncovering the structures that may be present in high dimensional data sets. For example, while many algorithms might get close, relatively few would be able to group the data as shown below (Figure 1.2).

Each clustering technique has its weaknesses and strengths, and these must be considered in conjunction with the goals of the analysis and the nature of the data. For example, many parametric clustering algorithms tend to find

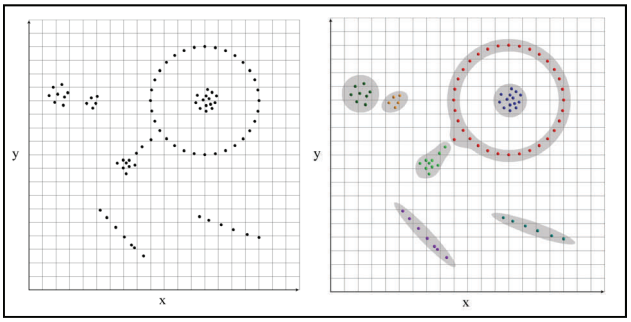


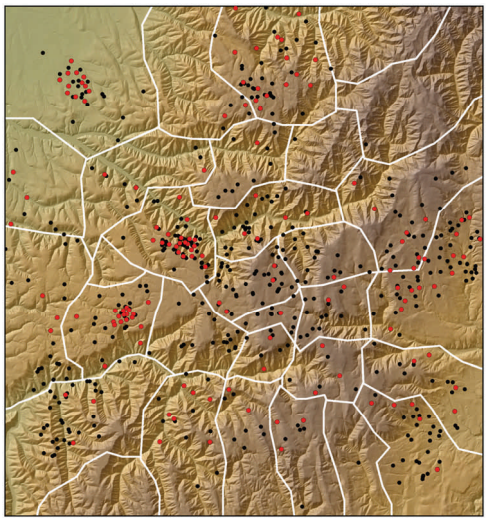
Figure 1.2. *Hypothetical cluster assignment of points (adapted from Jain et al. 1999).*

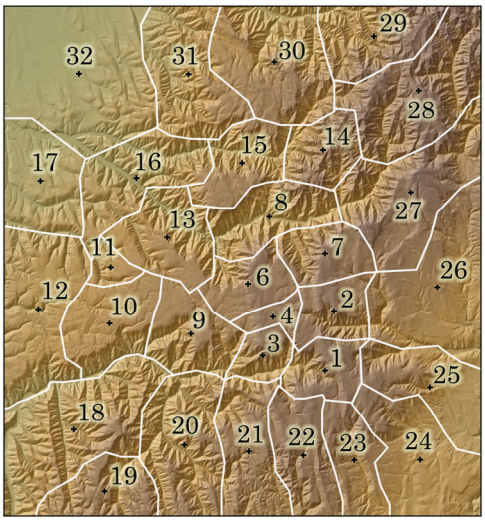
clusters of a particular shape (e.g., spherical), or of equal variance. However, CWD surveillance data from harvested deer are often distributed in elongated patterns (e.g., following a drainage or roadway, or within a valley.) and are not likely to yield clusters of equal variance. The spatial irregularity of CWD surveillance data reduces the choice of algorithms to those able to detect irregularly shaped clusters. Determining the distributional nature of the data set and selecting an appropriate clustering algorithm can be a time consuming and confusing process. It is important to remember that there is no "best" method covering every

situation. Accordingly, understanding strengths and weaknesses of each method is an important part of the clustering process. For example, clustering methods such as *k*-means and Ward's minimum variance method (i.e., least squares criterion algorithms) tend to select clusters with roughly the same number of observations in each cluster. Algorithms based on nonparametric density estimation such as single linkage and density linkage are generally considered to be the least biased methods for selecting clusters (SAS Institute Inc., SAS OnlineDoc® 9.1.3, Cary, NC: SAS Institute Inc., 2002-2005). However, this comes at the cost of reduced power to detect clusters. The specific goals of the analysis along with the knowledge of researchers and managers involved will be integral to the selection of appropriate clustering methods.

Clustering analysis at the regional scale can be used to determine the location of clusters as well as to answer questions about their significance (Kulldorff and Nagarwalla 1995). For the clustering of disease data it is necessary to compensate for the uneven distribution of the sampled data (Kulldorff 1997). This is particularly true with CWD surveillance efforts which vary widely across larger spatial scales. We advocate the use of spatial scan statistics for the clustering of disease data and cover this method in greater detail as the focal approach in the next section on landscape level methods.

While the use of scan statistics to detect clusters at the regional scale follows the methods outlined in the next section (i.e. at landscape level), there is one important difference: the analytical unit used for spatial clustering. For CWD data across large regions, it is often the case that information about the exact location of all samples is missing. For example, hunter harvested deer, which contribute overwhelmingly to CWD data sets, often lack geographic coordinates, but contain courser spatial information (e.g., wildlife management unit, county, etc). To make the data spatially explicit the analytical unit is shifted from the individual points to predetermined polygons. In other words, geographic coordinates (e.g., UTMs) are not required for individual CWD surveillance samples and data are effectively transformed into polygon count data (the number of positive and negative cases) with measures being concentrated at the central coordinates of the polygons (Figure 1.3). Polygons can take many forms including agency units (e.g., wildlife or game management units), administrative units (e.g., counties or postal codes), environmental units (e.g., water-sheds),





ID	negative cases	positive cases	X coordinate	Y coordinate
1	4	3	419191	5064761
2	13	2	419914	5069496
3	6	2	414249	5065942
4	10	1	415066	5069035
6	35	4	413084	5071668
7	19	1	419202	5074073
8	26	6	414767	5077020
9	18	0	408526	5067691
10	22	18	402052	5068516
11	0	0	402135	5073003
12	19	3	396383	5069635
13	44	21	406629	5075373
14	1	0	419037	5082333
15	10	1	412608	5081329
16	7	3	404192	5080139
17	6	3	396564	5079901
18	33	3	399199	5060031
19	0	0	401646	5055115
20	10	7	408010	5058794
21	7	2	413160	5058271
22	5	3	417485	5057961
23	1	0	421529	5057591
24	25	5	426775	5057615
25	12	4	427561	5063413
26	45	15	428152	5071399
27	23	8	426043	5078996
28	1	0	426657	5087127
29	0	0	423096	5091425
30	30	16	415164	5089414
31	4	2	408335	5088443
32	19	12	399611	5088477

Figure 1.3. *(left) A set of hypothetical polygon "units" overlaid on a set of positive (red) and negative (black) cases. (middle) The centroid locations (black crosses) and id numbers for each of the 32 polygon units. (right) Associated table with counts of positive and negative cases along with centroid UTM coordinates.*

or even arbitrary grid cells or quadrats. Abundant digital GIS data permits wide discretion in the choice of criteria for selection spatial units.

We note that aggregating into polygons, such as wildlife management units, may be the only reasonable choice if a most of the surveillance samples lack geographic coordinates. The primary disadvantage of "collapsing" data from points to polygons is loss of spatial heterogeneity within the polygons. For example, the most basic polygon attributes will contain only spatial information on centroids (or boundaries) and counts of the number of cases and non-cases within each polygon (Figure 1.3). If such "global," or 'first-order' clustering methods mask significant variation within polygons, the analysis should be supplemented with additional 'second-order,' or local, results based smaller spatial units, or even the points themselves (if geographic data are available).

Questions addressed / model predictions:

1. *Identifies high prevalence disease areas.*
2. *Identifies regional-scale spatial disease pattern.*
3. *Identifies potential spatial covariates to disease pattern.*

Data required:

1. The polygon, which could be a large area for a regional analysis (e.g., wildlife management unit or county), from which each positive and negative sample was taken.

Output:

1. Assigns positive cases or areas to a particular cluster.
2. Maps of disease clusters.

General usefulness:

At the regional scale, the main strength of cluster analysis is that relatively course resolution data can be used to identify areas of high disease prevalence (at which management intervention could be targeted). Cluster analyses are exploratory, but can be useful for hypothesis generation. Cluster analysis is a valuable initial step in examining the spatial epidemiology of a disease.

Usefulness to CWD modeling and/or management:

Cluster analysis is a valuable descriptive tool for CWD surveillance data. In general, we recommend that cluster analysis be conducted as a first step in the examination and evaluation of large-scale CWD surveillance data. The usefulness of cluster methods for CWD is the same as described above in "General usefulness." All CWD surveillance data, georeferenced to point or area, can be used in cluster analysis. This characteristic makes cluster analysis especially viable for multi-state data where some states collect sample coordinates and others do not.

If data are not collapsed into polygons for regional-scale applications, then all the spatial aspects of location-based cluster analysis need to be addressed. In the next section on landscape-level modeling, we discuss issues relevant to location-based cluster analysis in detail. In the section on Risk Analysis/Assessment, we consider the use of kernel density estimators, a special case of cluster analysis, to generate a risk surface.

SNAPSHOT APPROACH:
(FOLLOWING KEELING ET AL. 2004)

The goal of the snapshot approach is to identify areas to which a disease is likely to spread and the rate of spread given a single snapshot of the locations and disease status of an area. This method is appealing because it only requires one data collection effort, which is less expensive and time-consuming than other approaches. While Keeling et al. (2004) refer to this as a lattice-based, grid-based, or a cellular automata approach, the applicable area could be a county or wildlife management area rather than a square grid cell. Pairwise status (infected:infected, infected:non-infected, or non-infected:non-infected) between each area

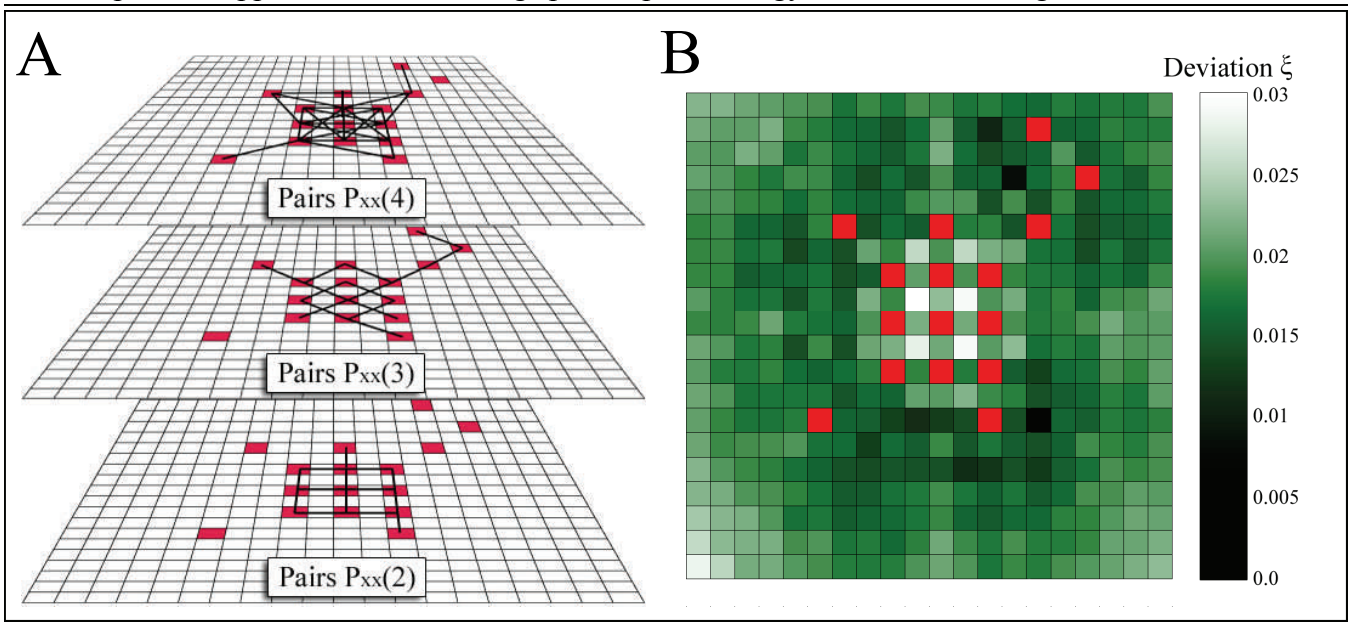


Figure 1.4. *(A) Pairwise status of infected:infected (P_{xx}) at distances of 2, 3, and 4 for a simple spatial distribution of positive CWD cases (red squares). Two cells joined by a black line represent a single pair at distance d. This pattern shows aggregation at intermediate distances and strong density dependence at d = 1 (because no red squares are adjacent to each other). (B) Likely sites of future CWD infection based on the spatial structure from (A) across all possible distances. Non-occupied sites are shaded (green) to indicate the deviation from the spatial structure in the observed snapshot. Dark squares represent the least change, and therefore represent sites that are most consistent with the observed spatial pattern. If the system is in statistical equilibrium (stationary prevalence), then these dark cells represent the most likely areas for positive cases of CWD. (Adapted from Keeling et al. 2004)*

or cell are calculated, along with the distances between areas. From the data, a function is fit to estimate how the risk of infection changes with distance from any positive cell. There are four main assumptions required for this method: 1) the spatial sample is a representative sample of the disease on the landscape, 2) the spatial area is homogeneous, 3) the spread is not limited by boundaries (e.g., roads, rivers, or other boundaries to animal movement), and 4) estimated parameters and mechanisms were constant during the formation of the spatial pattern. If these assumptions are met, the rate of change in the pairs and the probability of areas to become positive are identifiable (Figure 1.4).

Questions addressed / model predictions:

1. Estimates rate of spread as a function of distance from an infected area.
2. Estimates the probability that an area will become infected.

Data required:

1. Disease status of spatial areas, such as wildlife management units (presence-absence).
2. Centroid coordinates of the spatial areas (distances are calculated between area centroids).
3. Meet the 4 assumptions listed above.

Output:

1. Estimates the probability that an uninfected area will become infected.
2. Estimates the rate of spread as a function of distance from an infected area.

General usefulness:

The snapshot model is probably most useful at very broad or very fine scales, where the spread of an infection may be independent of temporary barriers. That is, at very broad

scales, many 'local' barriers are not of great interest or importance to modeling probability of infection. At very fine scales, areas within or between barriers can be modeled. The main strength of the snapshot method is that data are required at only one point in time, and the data are relatively straightforward to collect. As a result, the snapshot method permits a rapid assessment of where the disease may spread to next.

Usefulness to CWD modeling and/or management:

The data required, disease status of an area, and its centroid, are easily determined from CWD surveillance data. However, some of the model assumptions may be difficult to meet. Surveillance sampling could be designed to meet the first assumption over large areas such as wildlife management units, but none of the remaining three model assumptions are realistic for CWD epidemiology. In particular, the patchy distribution of hosts across the landscape is likely to be a strong confounding factor in a snapshot analysis. Given the restrictive assumptions, the snapshot approach is unlikely to be useful for modeling and management of CWD. However, the idea of using area-based, presence-absence disease data to model the probability of disease in an area (even when disease is not detected) is explored in more detail below.

OCCUPANCY ANALYSIS:

If presence-absence data are collected and modeled within a mark-resight framework (MacKenzie et al. 2002, MacKenzie et al. 2003), they can be used to estimate and monitor occupancy, colonization, or extinction probabilities of a wildlife species in a given area. In this case, occupancy is a proxy for abundance in large-scale monitoring studies. Data collection for occupancy estimation has the added advantage of requiring a less intensive field protocol and being potentially less costly as compared to methods for density or abundance estimation. Recent theoretical advances in the development of occupancy models and their implementation have dramatically increased the viability of using this technique for landscape-scale modeling (McKenzie et al. 2002, McKenzie et al. 2005, Royle and Dorazio 2006. Freeman et al. 2007).

CWD data could be modeled using occupancy modeling. Probability of disease occupancy in an area, such as a wildlife management unit, could be estimated and modeled similarly to occupancy of a species in an area. Because CWD does not appear to spread rapidly, one could assume that over a short time frame, colonization and extinction probabilities are sufficiently low to be irrelevant (although these may be the parameters of interest for diseases with fast-moving epidemic fronts). If one controls for sampling intensity (subsampling could be used to do this) and prevalence is relatively constant over time, then occupancy probabilities could be used as a proxy for prevalence probabilities over the same area. The disadvantage of this method, however, is a great loss of resolution. That is, prevalence could drop dramatically, but this drop would not be detected by an occupancy model. Its main use would be for very large areas, such as multiple states, where it could be adapted for areas with sparse sampling and different sampling protocols.

Similar to occupancy for monitoring wildlife species, data collection would be far simpler and less costly to collect compared to data collected to generate precise prevalence estimates or a risk surface. This approach readily lends itself to a model selection framework, similar to the risk assessment approach. Hypothesized environmental and ecological risk factors could be evaluated with respect to the probability of disease occurrence. Occupancy models would be especially tractable for very broad scales, although it may be an important challenge to correctly identify the size of the areas sampled for disease "occupancy". The area should be large enough to be efficiently sampled, but small enough to capture the spa-

tial heterogeneity of CWD prevalence (e.g., if you sample a large enough area, all areas will be positive) and to reflect the epidemiology of the disease and the use of space by hosts.

Questions addressed / model predictions:

1. Over a very broad scale, identifies areas for which the disease is present or absent.
2. Identifies areas having a low probability of disease detection due to sampling schemes.
3. Identifies environmental/ecological (abiotic and biotic) factors associated with the probability that the disease is present.
4. Identifies environmental/ecological (abiotic and biotic) factors/covariates associated with the probability that the disease is detected

Data required:

1. Disease status (presence or absence) of an area such as wildlife management unit.
2. At least 2 years of data for each area.

Output:

1. Estimates probability of disease presence or detection in an area, such as a wildlife management unit or county.
2. Estimates probability surface of disease occupancy using centroid of a given area.
3. Estimates effect sizes (e.g., difference in disease presence between treatment and control areas, between species, etc.), and other relevant statistics for factors in model.
4. If spatial environmental/ecological (abiotic and biotic) factor/covariate data were collected, then estimates their importance to presence and detection.
5. Provides model selection statistics.

General usefulness:

If designed correctly, an occupancy approach is, potentially, a cost-efficient method to monitor status of disease presence over very broad spatial scales. Methods to determine required sample sizes for predicted disease detection probabilities and disease prevalence have been described (Samuel et al. 2006). The method appears viable for highly infectious, fast-spreading diseases, as well as non-infectious or slow-spreading diseases. Also, the binomial and multinomial mark-resight methods underlying the estimation of occupancy probabilities explicitly estimate and accounts for spatial covariance (McKenzie et al. 2005). The occupancy approach is attractive because it is simple compared to other methods that model spatial correlation.

Similar to risk assessment, occupancy models would be more applicable for diseases with short dormancy/latency periods than for those with long latency periods. Environmental or ecological factors associated with short latency period, would be more readily identified because longer latencies may dilute the effect as evaluated in future time periods unless these factors were relatively constant and migration rates relatively low (e.g., animal stays in the place where the risk occurs).

Usefulness to CWD modeling and/or management:

Although occupancy models have not been used for disease modeling, they appear potentially useful for initial modeling of CWD at regional or other broad scales. Because relatively little data is required to determine CWD 'occupancy' status (i.e., is there one or more infected animals in an area?), most states have comparable data. Differences in data collection protocols can make data incomparable and hamper prevalence estimation and use of other approaches (e.g., cluster analysis and risk analysis, based on risk surface). The main drawback of occupancy modeling is a reduction in the resolution of biological inferences. That is, the output is probability that the disease is present, but it could be present at a very low prevalence or very high prevalence. This reduced resolution in the output would dilute the ability to realistically evaluate spatial factors and covariates. Again, choosing the correct spatial area over which to estimate disease presence or absence would be critical to reduce the loss of resolution.

EPIDEMIC TREES – A NETWORK ANALYSIS (FOLLOWING HAYDON ET AL. 2002)

Epidemic trees allow estimation of two primary values, R_0 and R_t. R_0 defines the average number of secondary cases that arise from a single case at the start of a disease outbreak or epidemic. R_t is the average number of secondary cases arising from a single infection during time = t. Traditional Susceptible-Infected-Recovered (SIR) models also estimate R_0 based on a theoretical model. The epidemic-tree approach is novel in that it is an empirical method of direct estimation of R_0 from the history of the observed cases (Haydon et al 2003). The epidemic tree approach is contingent upon data that accurately tracks the historical progression of the disease, specifically the temporal path from initial case to subsequent cases.

Questions addressed / model predictions:

1. Estimates disease transmission rate (R_0 and R_t) through time and space.
2. Evaluates, retrospectively, the effectiveness of different control strategies; i.e., estimates reduction of R_t for a control strategy instituted at a given location and time.
3. Allows retrospective comparisons, sensitivity, and cost-benefit analyses of different control measures, different timing of control actions (relative to onset of disease or relative to season), and different control locations.
4. Evaluates the influence of long-range transmission events compared to short-range transmission.
5. Identifies the best strategies available for a future outbreak or outbreak of similar disease.

Data required (* indicates data not currently collected as part of any CWD surveillance program):

1. Location of infected/infection.
2. Putative date of start of infection.*
3. Date a suspected infection reported.*
4. Date when infection confirmed.*
5. Viable rule set that ties subsequent cases to the case from which they originated.*

Output:

1. Estimates of R_t and its variance for specific time intervals and locations (estimates of rate of spread in time and space).
2. Estimates of generation time = interval between infection and subsequent case arising from it.
3. Estimates of reporting time = time between infection and subsequent reporting of case arising from it.
4. Estimates of routes of spread.
5. Estimates of rate of spread.

General usefulness:

Epidemic tree modeling is good for highly infectious, fast spreading epidemics. This method will underestimate R_t if all cases are not identified. However, as long as R_t is not biased by area (e.g., bias could arise if fewer infections identified away from urban areas due to different detection probability, whereas all areas would be equally under-represented a relatively non-biased situation), this method is usable for comparing the effect of different management strategies on R_t. Finally, for fast-spreading diseases that may re-invade, the strategies gleaned from retrospective analysis could be applied to future invasions.

Usefulness to CWD modeling and/or management:

The epidemic tree approach is not useful for modeling spatial epidemiology of CWD because positive animals cannot be linked to an originating case. Because CWD may be transmitted indirectly through a prion-contaminated environment, there may be no specific originating case or location. Also, it would be difficult to estimate a time of infection, since little is known about the course of disease in free-ranging deer, or about potential individual variation in disease progression. In addition, the timescale of CWD may be too long to take timely advantage of retrospective strategies.

BROAD SCALE FOCAL APPROACHES:

For the regional scale, we decided to present 2 focal approaches, micromaps and risk analysis. We present micromaps as a useful visualization technique to display data over large regions because they integrate spatial and temporal aspects of data. This technique would be the first step to summarize and examine regional data, while risk analysis is a logical second step. Risk analysis, at the regional scale, focuses on the two main questions: 1.) What are the significant CWD risk factors for a free-ranging mule deer population? and, 2.) Can we predict where CWD is likely found or will spread next? We focus on mule deer for all focal approaches, but these approaches could be applied to white-tailed deer and elk. Note: we do not use real data or perform a detailed analysis; our goal is simply to provide a general overview and illustrate the potential of these methods.

FOCAL APPROACH #1: LINKED MICROMAPS

Step #1- Generating Linked Micromap Plots:

After delineating game management unit (GMU) boundaries, linked micromap plots for the GMUs and the state of Colorado were created using the S-plus statistical software package. The sample S-plus code for creating linked micromap plots is available at Dan Carr's ftp site (ftp://galaxy.gmu.edu/pub/dcarr/newsletter/micromap/).

For CWD in northcentral Colorado, we simulated a sample that was limited to only a few GMUs. Therefore, micromap visualization can consist only of those GMUs with sampled deer. Prevalence and associated confidence intervals were subsequently plotted in a linked micromap representation.

Step #2- Interpreting the Final Linked Micromap Plots:

Figure 1.5 shows a series of linked micromap plots for the GMUs sampled for CWD. In the first linked micromap (1.5.A), four vertical panels (columns) are linked by geographic location, which is the GMU in this case. The map (panel 1) shows the boundaries for GMUs inside a CWD endemic area. The legend to the right of the map (panel 2) shows the GMUs designation with a dot in the linking color. The set of four graphs (panels 3 and 4) illustrate two statistical variables. In this particular example, dot-plots represent prevalence for the 2

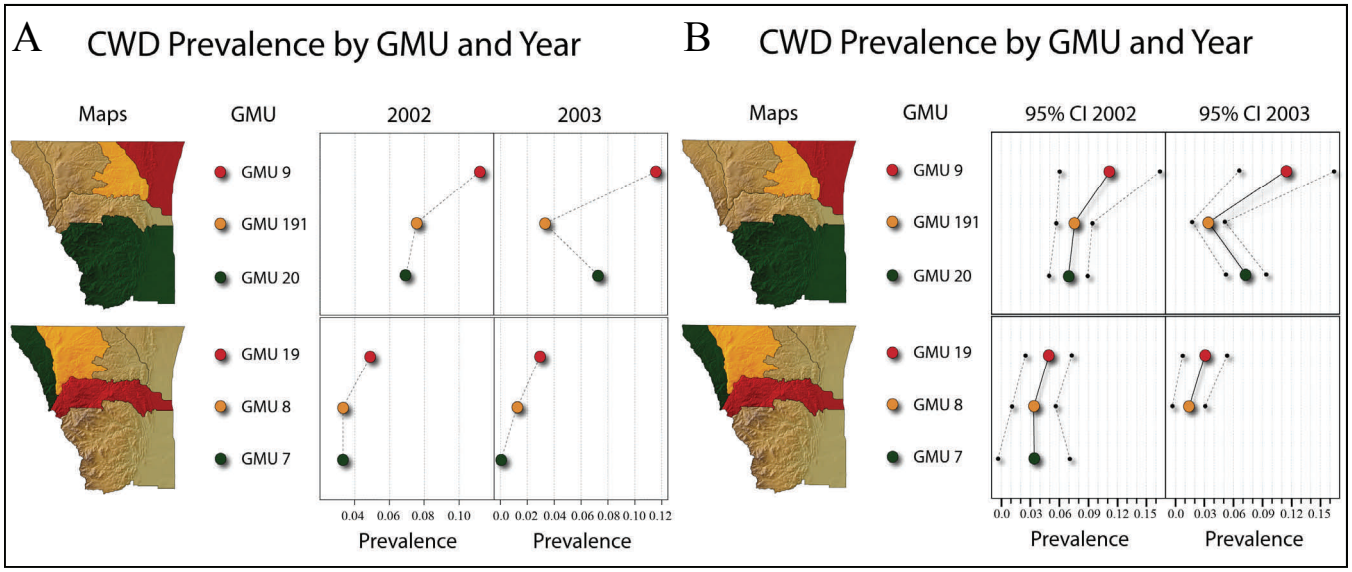


Figure 1.5. *Various micromaps for CWD from simulated mule deer data for northcentral Colorado. (A) CWD prevalence in 6 GMUs for 2002 and 2003. (B) An example of portraying prevalence across space (6 GMUs) and time (2 years) with 95% confidence intervals.* ***Note: this is not real data.*** *See text for more detailed explanation*

years in question, i.e., 2002 (panel 3) and 2003 (panel 4). All corresponding micromaps, labels, and statistical panels are linked by their colors. Note that three distinct colors are used to distinguish the GMUs within a particular micromap frame and are unique to that micromap. The GMUs are ranked according to simulated prevalence of 2002 from highest to lowest and are partitioned into two micromaps. The ranking is user-defined and could be 2003 if desired, while the partitioning and number of micromaps are based on the number of geographical units to be represented (see Symanzik and Carr (2007) for details). While offering no interpretation of the CWD data, it is immediately obvious which GMU had the highest and lowest prevalence in 2002. The GMUs in the lower micromap panel show a decline in prevalence from 2002 to 2003, but still maintain the same ranking. GMU9 was markedly static between years with the highest prevalence.

Step #3- Displaying and Interpreting Supplemental Statistical Information:

A further capability (i.e., supplementary statistical representation) of micromaps, displayed in Figure 1.5.B, is the addition of confidence intervals as a component of the prevalence panel. The confidence intervals (panels 3 and 4) represent the 95% lower and upper confidence limits. The larger colored dots refer to, as before, the prevalence in each GMU. One can now appreciate the fact that the prevalence of each GMU are not quite the "true" (actual) prevalence and that the confidence intervals describe uncertainties of the estimates. Moreover, readers can also observe that GMUs where simulated prevalence was estimated from limited data versus ample data as signified by the width of the confidence interval. As an example, consider how GMU9, which had the highest prevalence in both years, compares to GMU191. Upon initial examination of the prevalence information, it appears that GMU9 has a higher prevalence than GMU191. However, GMU9 has a wider confidence interval indicating that the prevalence for GMU9 is less reliable.

FOCAL APPROACH #2: RISK ANALYSIS USING PREVALENCE (OR ODDS RATIO) SURFACE

Step #1- Development of GIS Database:

Geographical risk modeling begins with the development of a GIS database that will be used to store, access, update, and model risk associated with a disease. Special attention should be placed on checking the quality, type, and spatial resolution of each data layer. Data quality is crucial to the modeling process and can have a significant impact on overall uncertainty. For example, the layer with the coarsest spatial resolution typically constrains the spatial resolution of the analysis to no less than that particular scale. For multi-state modeling, data for each risk factor must be standardized across the entire region. Because data often come from different sources, standardization may require reprocessing data to a common coordinate system or reclassifying categorical variables. This is an important process and maps of an individual risk factor can generate useful insights. These maps can help identify influential variables and/or areas that are problematic to overall risk analysis.

Examples of Potential Risk Factors:

- Proximity to known CWD positive deer and/or elk based on surveillance locations.
- Proximity to captive deer and/or elk facilities (e.g., farms or hunting ranches).
- Proximity to deer and/or elk feeding or wintering areas.
- Proximity to deer and/or elk processing facilities.
- Proximity to deer and/or elk research facilities.
- Proximity to taxidermy operations.
- Historical intensity of sheep grazing and current sheep density.
- Black-faced sheep density.
- Proximity to known scrapie infected or exposed sheep flocks.
- Deer and/or elk movement corridors.

Step #2 – Transforming the Data:

Variables usually require transformation to

a common format. GIS data can be stored as points, lines, polygons, or in raster format (2 dimensional arrays where each cell, or pixel, contains a single values). Raster data is well suited to modeling exercises such as risk analysis (Longley et al 2002). However, "rasterizing" data results in homogeneity within pixels (i.e., each cell is assigned a single value); therefore it is important to consider potential interactions between resolution and the nature of the data being rasterized.

In the case of risk analysis, the ultimate goal is to generate a value that represents the risk in any given cell as a function of the risk variables. This often requires restructuring the data into more biologically meaningful terms. For example, captive cervid facilities are a known potential source of CWD exposure risk. Locations of these facilities are usually provided in the form of polygons or points, and each location will be represented as either present or absent when transformed into a raster format. For each cell containing a facility, the entire cell is assigned a 'present' value, regardless of the size of the facility relative to the cell (e.g., the facility might be 5 acres, but the cell size could be 5x5 miles, or 16,000 acres – the entire 16,000 acres would be considered occupied by the facility). The real value of interest is how far any particular cell is from the nearest game farm, and the estimated distance may depend substantially on the process used to standardize the underlying data. In cases where distance is the parameter of interest, the surface data for that variable is expressed in terms of distance (Figure 1.6).

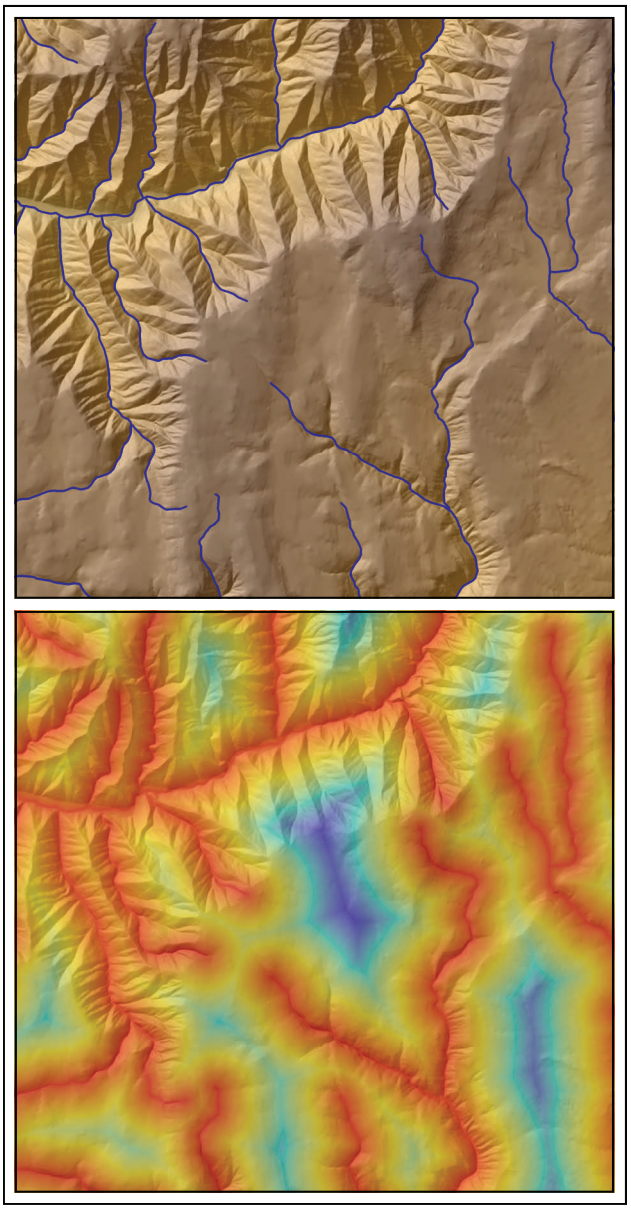


***Figure 1.6**. Transformation of rivers (top) from a linear feature to a raster data set (bottom). Each cell in the raster contains a value of the linear distance to the nearest river. The distance values increase as the color changes from red to yellow to blue.*

Step #3 – Assigning Risk:

Risk is assigned to individual cells for each risk layer. The output is the result of applying a "risk" equation to each pixel, where each layer is a variable and the relative risk factors are coefficients between 0 and 1. The final map illustrates results from the equation. Figure 1.7 shows a visualization for two variables which are considered to be important factors in the spread of CWD: proximity to existing cases of CWD and proximity to captive cervid facilities. Note that for a real data set, risk would be assigned to the predictor variables based on their importance in a spatially corrected regression analysis. Techniques such as hierarchical partitioning (Chevan and Sutherland 1991), partial correlations (Cox 1985), and other measures such as variable importance in projection (Birkner and van der Laan 2006) can be used to determine variable weighting. Once the variables are ranked, the risk model can be

used to forecast risk for areas where disease data have not been collected. Where data on a predictor variable are not available, expert judgment may be used to assign relative variable importance, with a corresponding reduction in confidence in the forecasts.

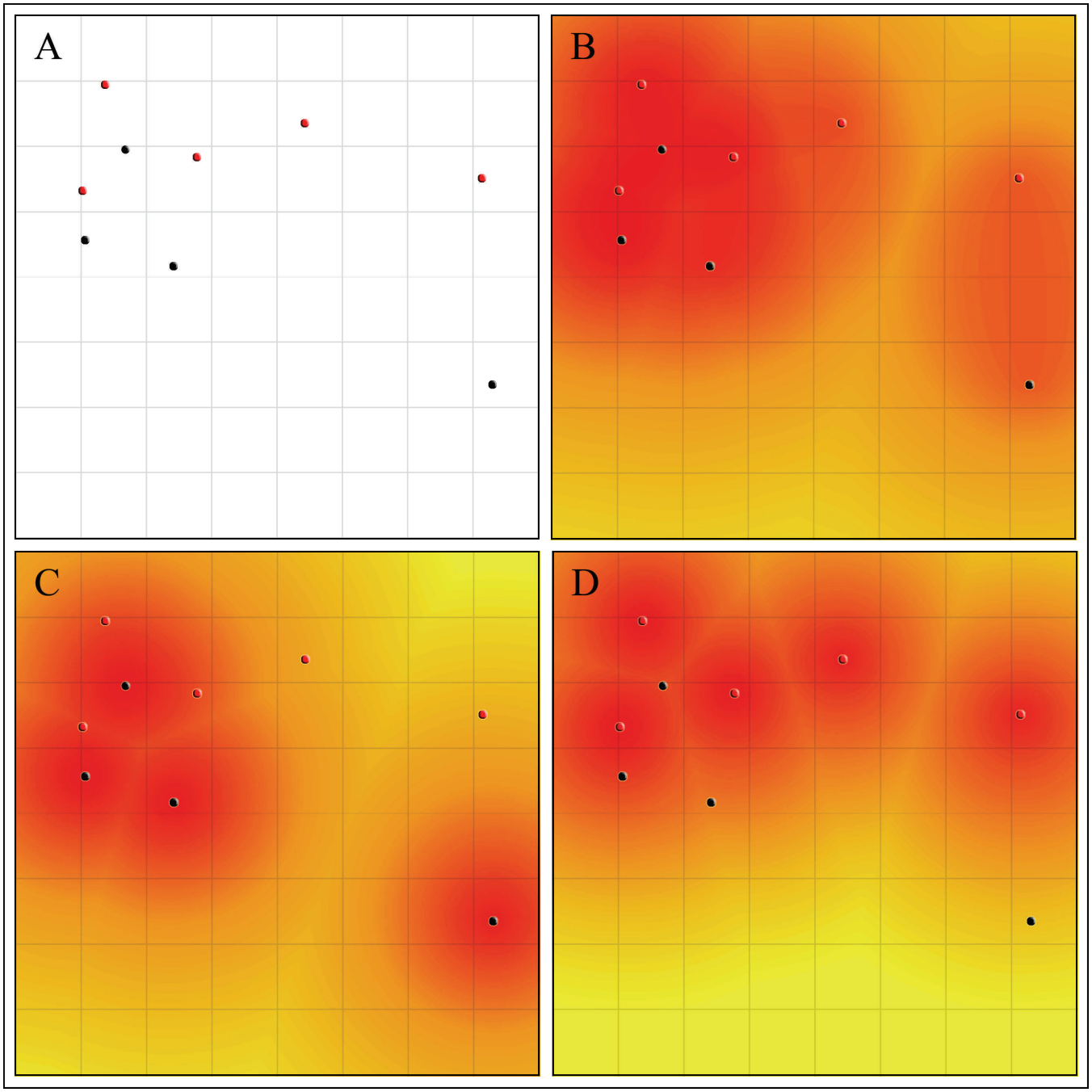


Figure 1.7. *(A) Hypothetical locations of two chronic wasting disease risk factors, positive chronic wasting disease cases (red points) and captive cervid facilities (black points). Risk surfaces when (B) risk is weighted equally between proximity to chronic wasting disease positives and proximity to captive cervid facilities, (C) risk from proximity to captive cervid facilities is 10x that of the risk from proximity to positive cases, and (D) risk from proximity to positive cases is 10x that of the risk from proximity to captive cervid facilities. In (B-D) red represents a relatively high level of risk and yellow represents a relatively low risk. The black grid represents large scale quadrats for visualization, while the surface is modeled at a much finer resolution. The surfaces in this figure were built from a 4000x4000 cell lattice.*

To highlight how important assigning relative risk weight to variables is to the overall risk analysis, we have provided three variations. In the first case, we assign equal risk weights to the proximity to existing cases of CWD (weight = 0.5) and proximity to captive cervid facilities (weight = 0.5). In the second case we assign 10 times the weight to the risk of proximity to captive cervid facilities (weight = 0.91) compared to proximity to existing CWD positives (weight = 0.09). In the final case, we reverse the weighting and assign 10 times the weight to the risk of proximity to existing CWD positives (weight = 0.91) compared to proximity to captive cervid facilities (weight = 0.09).

Step #4 – Displaying Risk Output:

Step 4 involves delimiting areas of relatively high and low risk based on the output layer. Typically isopleths of predetermined levels are shown in different colors allowing for a visual representation of the risk involved (See Figure 1.7). However, threshold levels can also be used to show areas above a set level of risk.

Step #5 – Model Validation:

Model validation is an important part of the risk assessment process. Both internal validation and external model validation should be assessed, especially if management decisions are based on risk models. Internal (e.g., bootstrapping) is the more straight forward procedure as the data already exist. This can be considered a self-consistency check, as any systematic differences between the simulations and the data (upon which the model is based) indicate weaknesses in the model (Gelman et al. 2004). Internal validations tend to be overly optimistic about model performance because the data for modeling and validation come from the same data set and they are thus not independent. External validation provides a more reliable estimate of sensitivity (proportion of false negatives) and specificity (proportion of false positives). External validation compares the fit of model predictions to new data (Gelman et al. 2004). For example, if CWD can be considered stationary and surveillance patterns/intensity are consistent across time, then it may be possible to use temporal external validation, such as building the model from one year of surveillance data and testing it on data from subsequent years. See Gelman et al. (2004) for a detailed explanation of model validation.

General Conclusions About Regional-Scale Approaches for CWD Modeling

Although contagious, CWD appears to fall somewhere between the class of slowly spreading diseases, such as rabies, and non-infectious chronic diseases, such as cancer or a pollution/toxin induced illness. CWD is similar to TB in that it spreads among individuals in a wildlife population by direct or indirect contact, and has a long latency period. Existing epidemiological models applied at a regional scale tend to fall into 2 categories: those more useful in representing infectious, relatively quick-spreading diseases or those more useful for non-contagious, spatially static diseases. We know little about the spread of CWD in the wild, and recently detected foci may result from increased surveillance sampling rather than spread. Consequently, methods for non-infectious diseases, such as cluster analysis and risk analysis, seem the most appropriate for spatial modeling and portrayal of CWD at a regional scale.

Thus, we chose risk analysis as the focal approach for large-scale modeling. Cluster analysis is a similarly useful method, and we describe use of cluster analysis as the focal approach at the landscape level in the next section. In general, we recommend micromaps as a first step for describing CWD and communicating patterns at broad scales. For many purposes, the next step is likely to be a risk analysis. The utility of snapshot or occupancy approaches for analyzing CWD spatial epidemiology is uncertain, and these approaches require further testing with CWD or similar diseases. We recommend that simulations (e.g., determining power to detect a specific change

in prevalence given different sample sizes) be conduced to evaluate their viability for CWD data at the regional scale. Finally, we note that epidemic trees are inappropriate for evaluating CWD spread because these were developed for analysis of highly infectious, fast spreading diseases.

The data gaps for these methods are primarily large spatial environmental and ecological data. These data often exist, but usually they need to be compiled and standardized across jurisdictional boundaries, for the region of the analysis. As noted in the focal approach, the quality, type, and spatial resolution of each data layer needs to be evaluated as part of the modeling process. The development of reliable, well-documented, spatial GIS-based layers of relevant biological and ecological factors will strongly promote CWD modeling efforts at the regional scale.

LITERATURE CITED:

Allan, B. F., F. Keesing, and R. S. Ostfeld. 2003. Effect of forest fragmentation on Lyme disease risk. Conservation Biology 17:267-272.

Anderson, D. P., J. D. Forester, M. G. Turner, J. L. Frair, E. H. Merrill, D. Fortin, J. S. Mao, and M. S. Boyce. 2006. Factors influencing female home range sizes in elk (*Cervus elaphus*) in North American landscapes. Landscape Ecology 20:257-271.

Birkner, M. D., and M. J. van der Laan. 2006. Application of a Variable Importance Measure Method," The International Journal of Biostatistics. 2: Article 6.

Chevan, A., and M. Sutherland. 1991. Hierarchical partitioning. The American Statistician. 45:90-96.

Cox, L. A. Jr. 1985. A new measure of attributable risk for public health applications. Management Science 31:800-813.

Briggs, D. J. 2000. Exposure assessment. Pages 335-359 *In* Spatial Epidemiology. P. Elliott, J. C. Wakefield, N. G. Best, and D. J. Briggs, Editors. Oxford University Press, Oxford, England.

Carr, D. B., and S. M. Pierson. 1996. Emphasizing statistical summaries and showing spatial context with micromaps. Statistical Computing and Statistical Graphics Newsletter. 7:16-23.

Carr D. B., A. R. Olsen, J. P. Courbois, S. M. Pierson , and D. A. Carr. 1998. Linked Micromap Plots: Named and Described. Statistical Computing and Statistical Graphics Newsletter 9: 24-32.

Carr D. B. 2001. Designing Linked Micromap Plots for States with Many Counties. Statistics in Medicine 20:1331-1339.

Dent, B. D. 1993. Cartography: Thematic Map Design (Third Edition). William C. Brown, Dubuque, Iowa, USA.

Dunford, W., and K. Freemark. 2005. Matrix matters: Effects of surrounding land uses on forest birds near Ottawa, Canada. Landscape Ecology 20:497-511.

Freeman, S. N., D. G. Noble, S. E. Newson, and S. R. Baillie. 2007. Modeling population changes using data from different surveys: the common birds census and the breeding bird survey. Bird Study 54:61-72.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 2004. Bayesian Data Analysis: Second Edition. CRC Press, Boca Raton, Florida, USA.

Harris, R. L. 1999. Information Graphics – A Comprehensive Illustrated Reference. Oxford University Press, New York, New York, USA.

Haydon, D. T., M. Chase-Topping, D. J. Shaw, L. Matthews, J. K. Friar, J. Wilesmith, and M. E. J. Woolhouse. 2003. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. Proceedings of the Royal Society of London B 270:121-127.

Hess, G. R., S. E. Randolph, P. Arneberg, C. Chemimi, C. Furlanello, J. Harwood, M. G. Roberts, and J. Swinton. 2002. Spatial aspects of disease dynamics. Pages 102-118 *In* The ecology of wildlife diseases. P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heesterbeek, and A. P. Dobson, Editors. Oxford University Press, Oxford, England.

Hickling, G. J. 2002. Dynamics of bovine tuberculosis in wild white-tailed deer in Michigan. Michigan Department of Natural Resources Wildlife Division Report No. 3363. Lansing, Michigan, USA. Available at: http://www.michigandnr.com/publications/pdfs/huntingwildlifehabitat/reports/deer_TB_dynamics.pdf (accessed December 2006).

Jain, A. K., M. N. Murty, and P. J. Flynn. 1999. Data Clustering: A Review. ACM Computing Surveys, Vol. 31, No. 3.

Jenkins, S. R., and W. G. Winkler. 1987. Descriptive epidemiology from an epizootic of raccoon rabies in the Middle Atlantic States, 1982-1983. American Journal of Epidemiology 126:429-437.

Johnson, C. J., K. E. Phillips, P. T. Schramm, D. McKenzie, J. M. Aiken, and J. A. Pedersen. 2006. Prions adhere to soil minerals and remain infectious. Plos Pathogens 2:296-302.

Kallèn, A., P. Arcuri, and J. D. Murry. 1985. A simple model for the spatial spread and control of rabies. Journal of Theoretical Biology 116:377-393.

Keeling, M. J., S. P. Brooks, and C. A. Gilligan. 2004. Using conservation of a pattern to estimate spatial

parameters from a single snapshot. Proceedings of the National Academy of Sciences 101:9155-9160.

Kitron, U. 2000. Risk maps: mapping transmission and burden of vector-borne diseases. Parasitology Today 16:324-325.

Kulldorff, M., and N. Nagarwalla. 1995. Spatial disease clusters: detection and inference. Statistics in Medicine 14:799-810.

Kulldorff, M. 1997 A spatial scan statistic. Communications in Statistics: Theory and Methods 26:1481-1496.

Longley, P.A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind. 2002. Geographic information systems and science. 3rd edition. John Wiley & Sons, Chichester, England.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248-2255.

MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and. A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84:2200-2207.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2005. Occupancy estimation and modeling. Academic Press, Burlington, Massachusetts, USA.

Merler, S., C. Furlanello, C. Chemini, and G. Nicolini. 1996. Classification tree methods for analysis of mesoscale distribution of *Ixodex ricinus* in Trentino, Italian Alps. Journal of Medical Entomology 33:888-893.

Moore, D. A. 1999. Spatial diffusion of raccoon rabies in Pennsylvania, USA. Preventative Veterinary Medicine 40:19-32.

Morris, S. E., and J. C. Wakefield. 2000. Assessment of disease risk in relation to a pre-specified source. Pages 153-184 *In* Spatial Epidemiology. P. Elliott, J. C. Wakefield, N. G. Best, and D. J. Briggs, Editors. Oxford University Press, Oxford, England.

Murwira, A., and A. K. Skidmore. 2005. The response of elephants to the spatial heterogeneity of vegetation in a Southern African agricultural landscape. Landscape Ecology 20:217-234.

Prins, H. H. T., and F. J. Weyerhauser. 1987. Epidemics in populations of wild ruminants – anthrax and impala, rinderpest and buffalo in Lake Manyora National Park, Tanzania. Oikos 49:28-38.

Rogers, D., S. I. Hay, and M. J. Packer. 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. Annals of Tropical Medicine and Parasitology 90:225-241.

Royle, J. A. and R. M. Dorazio. 2006. Hierarchical models of animal abundance and occurrence. Journal of Agricultural Biological and Environmental Statistics 11:249-263.

Samuel, M. D., D. O. Joly, M. A. Wild, S. D. Wright, D. L. Otis, R. W. Werge, and M. W. Miller. 2003. Surveillance strategies for detecting chronic wasting disease in free-ranging deer and elk. USGS-National Wildlife Heath Center Report. Madison, Wisconsin, USA. Available at: http://www.nwhc.usgs.gov/publications/fact_sheets/pdfs/cwd/CWD_Surveillance_Strategies.pdf .

SAS Institute Inc., SAS OnlineDoc® 9.1.3, Cary, North Carolina, USA: SAS Institute Inc., 2002-2005.

Schauber, E. M., R. S. Ostfeld, and A. S. Evans Jr. 2005. What is the best predictor of annual Lyme disease incidence: weather, mice, or acorns. Ecological Applications 15:575-586.

Schmitt, S. M., S. D. Fitzgerald, T. M. Cooley, D. S. Bruning-Fann, L. Sullivan, D. Berry, T. Carlson, R. B. Minnis, J. B. Payeur, and J. Sikarskie. 1997. Bovine tuberculosis in free-ranging white-tailed deer from Michigan. Journal of Wildlife Diseases 33:749-758.

Sleeman, J. M. 2005. Disease risk assessment in African great apes using geographic information systems. EcoHealth 2:222-227.

Smith, G. C., and S. Harris. 1991. Rabies in urban foxes (*Vulpes vulpes*) in Britain: the use of a spatial stochastic simulation model to examine the pattern of spread and evaluate the efficacy of different control regimes. Proceedings of the Royal Society of London B 334:459-479.

Symanzik, J. and Carr, D. B. 2007. Interactive Linked Micromap Plots for the Display of Geographically Referenced Statistical Data. In, Handbook of Data Visualization. Chen, Chun-houh; Härdle, Wolfgang; Unwin, Anthony (Eds.). Springer. ISBN-10:3-540-33036-4.